Relative Weighting of Semantic and Syntactic Cues in Native and Non-Native Listeners' Recognition of English Sentences

1897

Lu-Feng Shi[1] and Laura L. Koenig[1,2]

Objective: Non-native listeners do not recognize English sentences as effectively as native listeners, especially in noise. It is not entirely clear to what extent such group differences arise from differences in relative weight of semantic versus syntactic cues. This study quantified the use and weighting of these contextual cues via Boothroyd and Nittrouer's *j* and *k* factors. The *j* represents the probability of recognizing sentences with or without context, whereas the *k* represents the degree to which context improves recognition performance.

Design: Four groups of 13 normal-hearing young adult listeners participated. One group consisted of native English monolingual (EMN) listeners, whereas the other three consisted of non-native listeners contrasting in their language dominance and first language: English-dominant Russian-English, Russian-dominant Russian-English, and Spanish-dominant Spanish-English bilinguals. All listeners were presented three sets of four-word sentences: high-predictability sentences included both semantic and syntactic cues, low-predictability sentences included syntactic cues only, and zero-predictability sentences included neither semantic nor syntactic cues. Sentences were presented at 65 dB SPL binaurally in the presence of speech-spectrum noise at +3 dB SNR. Listeners orally repeated each sentence and recognition was calculated for individual words as well as the sentence as a whole.

Results: Comparable *j* values across groups for high-predictability, low-predictability, and zero-predictability sentences suggested that all listeners, native and non-native, utilized contextual cues to recognize English sentences. Analysis of the *k* factor indicated that non-native listeners took advantage of syntax as effectively as EMN listeners. However, only English-dominant bilinguals utilized semantics to the same extent as EMN listeners; semantics did not provide a significant benefit for the two non-English-dominant groups. When combined, semantics and syntax benefitted EMN listeners significantly more than all three non-native groups of listeners.

Conclusions: Language background influenced the use and weighting of semantic and syntactic cues in a complex manner. A native language advantage existed in the effective use of both cues combined. A language-dominance effect was seen in the use of semantics. No first-language effect was present for the use of either or both cues. For all non-native listeners, syntax contributed significantly more to sentence recognition than semantics, possibly due to the fact that semantics develops more gradually than syntax in second-language acquisition. The present study provides evidence that Boothroyd and Nittrouer's *j* and *k* factors can be successfully used to quantify the effectiveness of contextual cue use in clinically relevant, linguistically diverse populations.

Key words: Non-native, Semantics, Sentence recognition, Syntax.

(Ear & Hearing 2016;37;424–433)

[1]Department of Communication Sciences and Disorders, Long Island University–Brooklyn Campus, Brooklyn, New York, USA; and [2]Haskins Laboratories, New Haven, Connecticut, USA.

INTRODUCTION

Clinical practitioners frequently encounter individuals from diverse language backgrounds. If a client performs poorly on a recognition test involving linguistic stimuli, one must ask how linguistic background might contribute to that result. However, practitioners are typically not familiar with work in linguistics and bilingualism. This study investigates differences in linguistic cue use between native and non-native listeners following the model laid out by Boothroyd and Nittrouer (1988) and explores the application of *j* and *k* factors in these two clinically relevant populations. In what follows, evidence is first summarized for non-native disadvantages in linguistic cue use. Boothroyd and Nittrouer's study and subsequent studies employing their methods are then reviewed. Finally, research questions for the present study are proposed.

Language Background and Use of Linguistic Cues

The degree to which a listener successfully recognizes sentences varies with many factors. Two widely-studied factors are contextual redundancies in the sentence (Kalikow et al. 1977; Boothroyd & Nittrouer 1988; Nittrouer & Boothroyd 1990; Bronkhorst et al. 1994; Versfeld et al. 2000) and the language background of the listener (e.g., Mayo et al. 1997; van Wijngaarden et al. 2004; Mattys et al. 2010; Shi 2010, 2012). Native listeners make use of contextual redundancies (or cues) at morphological (Cutler et al. 1986), semantic (Benoit 1990; Bell & Wilson 2001; Aydelott & Bates 2004; Golestani et al. 2009), syntactic (Speaks & Jerger 1965; Nakatani & Dukes 1973), and prosodic (Akker & Cutler 2003; Mattys et al. 2005) levels. For example, Kalikow et al. (1977) found that native English listeners achieved perfect recognition when sentences were rich in context (mainly semantic cues), despite the presence of concurrent multitalker babble noise (+5 dB SNR and above). At the same SNR, performance was 20 to 30% lower when contextual cues were withheld. This finding suggests that bottom–up processing that relies on acoustic-phonetic information is not sufficient for error-free recognition in noisy environments.

Thus, top–down processes that utilize context help the listener fill in what is missing, based on her/his knowledge of the language, especially when the incoming stimulus is degraded (see McClelland et al. 2006 for a review). Non-native listeners are not as language proficient as native listeners and their use of these top–down processes is not as effective (e.g., Mayo et al. 1997; Akker & Cutler 2003; Clahsen & Felser 2006; Golestani et al. 2009; Mattys et al. 2010; Shi 2010, 2012). For instance, Mattys et al. compared how native British English speakers and Cantonese learners of English weighted acoustic-phonetic versus lexical-semantic cues in a speech segmentation task. When two words were presented as a string in noise, native listeners

0196/0202/2016/374-0424/0 • Ear & Hearing • Copyright © 2016 Wolters Kluwer Health, Inc. All rights reserved • Printed in the U.S.A.

<zdoi; 10.1097/AUD.0000000000000271>

tended to shift the weight to lexical-semantic cues and segment the string into meaningful words. By contrast, non-native listeners consistently relied on acoustic-phonetic cues and as a result their segmentations did not always result in meaningful words. These findings are taken as evidence that non-native listeners rely more on an acoustic-phonetically based bottom–up strategy to process speech.

Similarly, Mayo et al. (1997) controlled the semantic predictability of sentence-final keywords. They found that non-native listeners were not as successful as native listeners at extracting these cues to correctly identify keywords when sentences were presented in multitalker babble. Using the same set of sentences processed in reverberation as well as babble, Shi (2010) confirmed the Mayo et al. finding (1997) and further demonstrated that even native bilinguals (who self-reported to have acquired both languages since birth) could be significantly less effective than their monolingual counterparts on the task.

Although the studies cited above demonstrate differences between non-native and native listeners, they do not allow concurrent estimation of the speech recognition gain due to multiple cues. To directly quantify how a particular linguistic cue benefits speech recognition in native versus non-native listeners, the present study employed the models proposed by Boothroyd and Nittrouer (1988).

The *j* and *k* Factors

Boothroyd and Nittrouer (1988) expressed linguistic redundancies by two factors, *j* and *k*. Factor *j* describes the relationship between the components and the whole (segments in a word or words in a sentence), whereas factor *k* describes the relationship between components with and without context. It is assumed that recognition of the whole speech stimulus (a word or sentence) is based on recognition of its components (phonemes or words). When components are independent of one another, the probability of recognizing the whole is the product of the probability of recognizing all components. That is,

$$p_w = p_p^j, \qquad (1)$$

where p_w is the probability of recognizing the whole, p_p is the probability of recognizing each component, and *j* is the number of components that are statistically independent of one another. Equation (1) can be rewritten to express *j* as the logarithmic ratio between the probability of recognizing a whole unit and its components:

$$j = \frac{\log_{10} p_w}{\log_{10} p_p}. \qquad (2)$$

When there is no redundancy in the speech stimulus, *j* equals the total number of components in the stimulus (i.e., all components are independent of one another). When such independence is violated due to phonotactic, lexical, semantic, and/or syntactic constraints, *j* decreases. To illustrate, a non-meaningful sentence consisting of four unrelated words has four components, thus a *j* of 4. If the sentence is organized according to lexical-semantic and morphosyntactic rules and conveys a coherent meaning, then each word narrows down the list of other possible words that the sentence may contain. For example, "The student…" indicates that the ensuing word is likely a verb in a third-person singular form or a relative clause that modifies the subject according to English morphosyntactic rules. It also suggests the possible nature of the utterance, which makes related content words such as "classroom," "school," "study," etc. more likely to appear next than words, such as "sky" or "earthworm." Due to these constraints, words are not independent of one another in a meaningful sentence, yielding a *j* lower than the number of words in the sentence (Boothroyd & Nittrouer 1988; Nittrouer & Boothroyd 1990; Olsen et al. 1997; Eisenberg et al. 2000; Benki 2003).

To tease apart lexical-semantic and morphosyntactic cues, Boothroyd and Nittrouer (1988) designed three sets of four-word sentences in which the predictability of words was manipulated. High-predictability (HP) sentences included both semantic and syntactic cues (e.g., "Tough guys sound mean"), low-predictability (LP) sentences included only syntactic cues (e.g., "Thin books look bright"), and zero-predictability (ZP) sentences included neither cue (e.g., "Sing his get throw"). Results showed that *j* systematically decreased in value from ZP (4.13) to LP (3.33) and then to HP (2.21) sentences.

Boothroyd and Nittrouer's (1988) second factor, *k*, was developed to quantify the use of lexical, semantic, and syntactic cues in word and sentence recognition. The recognition probability of a stimulus can be expressed using probabilities of recognizing the components making up the stimulus. That is,

$$1 - p_c = (1 - p_i)^k, \qquad (3)$$

where p_c is the probability of recognition of the stimulus with context and p_i is the probability of recognition without context (i.e., none of the words in the sentence or phonemes in the word are connected). Based on Equation (3), *k* can be expressed as the logarithmic ratio

$$k = \frac{\log_{10}(1 - p_c)}{\log_{10}(1 - p_i)}. \qquad (4)$$

In sentences, *k* reveals the relative weights of different contextual cues. Three values for *k* were computed by Boothroyd and Nittrouer (1988). Because LP sentences contain only syntactic cues and ZP sentences no cues, *k* as the logarithmic quotient between these two types of sentences indicates the contribution of syntactic cues (i.e., k_s) to correct recognition of the whole sentence. In the same vein, *k* as the ratio between HP and LP sentences indicates the contribution of semantic cues (i.e., k_m), and *k* between HP and ZP sentences indicates the combined contribution of semantic plus syntactic cues (i.e., k_{s+m}). Boothroyd and Nittrouer obtained a higher k_m (1.97) than k_s (1.32) in native listeners, suggesting greater weight of semantics than syntax in sentence recognition.

As discussed by Boothroyd and Nittrouer (1988), the *j* and *k* factors yield some benefits relative to other recognition metrics. First, the nonlinear nature of the performance function means that a 10% increase in performance in one part of the scale (e.g., from 50 to 60%) is not equivalent to such a change at the margins of the scale (e.g., from 90 to 100%). A caution to this advantage is that, mathematically, extreme values (e.g., >95% or <5%) in raw performance could lead to unreliable *j* and *k* factors. Second, *j* and *k* factors provide a direct measure of contribution of partial recognition to the recognition of the entire stimulus. For example, *j* tells us how many words are needed to correctly pass on the information contained in the sentence. Raw performance may reveal a 20% difference between recognition of the whole sentence and recognition of each word in the sentence, but such a 20% difference does not inform us the power of redundancy across the words in the same sentence. How effectively listeners utilize these redundancies is of great interest not only to linguists but also educators and clinical practitioners. Therefore, this study explored *j* and *k* factors as advantageous measures that may be potentially used in clinical settings.

Purposes of the Present Study

The roles of *j* and *k* factors in sentence recognition have been evaluated in native listeners differing in age and hearing status (Boothroyd & Nittrouer 1988; Nittrouer & Boothroyd 1990; Olsen et al. 1997; Eisenberg et al. 2000), but not so much in non-native listeners. Of the few studies that have evaluated *j* and *k* factors in native versus non-native listeners, Nittrouer and Lowenstein (2010) obtained *j* for HP sentences, and Shi (2014) calculated *k* between HP and LP sentences. Both studies were thus focused on semantics, making it impossible to compare the relative weighting of semantics versus syntax as two essential contextual cues for sentence recognition. Furthermore, the above two studies were limited in the background of their non-native participants. Nittrouer and Lowenstein included 20 native Mandarin listeners who learned English as a second language beyond 12 years of age and were, at the time of the experiment, working in an academic setting (e.g., graduate student, postdoctoral fellow, or faculty). Shi included 33 non-native listeners with various first languages and ages of English acquisition varied widely (6 to 42 years old). As a result, the homogeneity of Nittrouer and Lowenstein's non-native listeners may restrict the generalizability of the findings, whereas the great diversity of non-native listeners included in Shi could have washed out potential effects of specific language background variables on semantic use.

Thus, given the limitations in the types of contexts and characteristics of non-native listeners in previous studies, the present study measured the extent of compromise at semantic and syntactic levels in non-native listeners' recognition of English sentences using Boothroyd and Nittrouer's *j* and *k* factors. Monolingual English listeners were compared with three groups of non-native listeners (native Russian listeners dominant in English, native Russian listeners dominant in Russian, and native Spanish listeners dominant in Spanish) to assess possible effects of language dominance and first language on *j* and *k*. Based on past work, the following predictions were made.

1. Factor *j* should show a pattern of HP < LP < ZP for all listeners. Factor *j* for all three types of sentences should be greater in non-native than native listeners, based on Nittrouer and Lowenstein's findings (2010). Due to differences in their language experience, perhaps the non-native listeners dominant in their first language would yield greater *j* than those dominant in English. It was also predicted that the intergroup differences could be greater for HP than the other two types of sentences since both semantic and syntactic knowledge contribute to recognition.
2. All listeners, native and non-native alike, should have a higher k_{s+m} than k_s or k_m because two cues combined should provide more information than either cue alone. Regarding listener groups, it was predicted that *k* would be lower in non-native than native listeners as the former's language skills might be still developing. Among non-native listeners, *k* would be lower in listeners dominant in their first language than in those dominant in English.
3. Specific prediction was difficult for the relative weight of semantic and syntactic cues in non-native listeners' sentence recognition given conflicting results from previous studies. Non-native listeners could have $k_s < k_m$, based on study by Sanders and colleagues (Sanders et al. 2002; Sanders & Neville 2003); that is, non-native listeners rely more heavily on semantic than syntactic cues. It is also possible that non-native listeners could have $k_s > k_m$, based on study by Mattys et al. (2010) and Shi (2014), both of which found that non-native listeners show deficits in semantic use. A third pattern would be $k_s = k_m$, indicating equal weighting of syntax and semantics.

MATERIALS AND METHODS

Participants

Data were collected from 52 adult listeners. All demonstrated normal hearing, with pure-tone thresholds no greater than 20 dB HL at octave frequencies 250 to 8000 Hz (American National Standards Institute 2010). These listeners were divided into 4 groups of 13 each, depending on their language background. English monolingual native (EMN) listeners (9 women, 4 men, 22 to 40 years old) reported that they were born and raised in a family where only English was spoken and they had never learned and used a language other than English beyond high school years.

All non-native listeners completed the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al. 2007). In their LEAP-Q responses, 13 of the 26 native Russian listeners self-identified as dominant in English (English-dominant Russian-English [RED], 7 women, 6 men, 21 to 36 years old). The other 13 indicated dominance in Russian (Russian-dominant Russian-English [RRD], 10 women, 3 men, 23 to 35 years old). The 13 native Spanish listeners all identified themselves as dominant in Spanish (Spanish-dominant Spanish-English [SSD], 10 women, 3 men, 21 to 40 years old). In combination, the RRD and SSD groups made up the non-English-dominant group. We mainly included Russian and Spanish listeners in this study because we have large and diverse Russian- and Spanish-speaking communities in Brooklyn and our sample reflected local demographics. The language profiles of the three non-native groups, obtained via the LEAP-Q, are summarized in Table 1.

We did not require non-native listeners to go through any objective measure of their dominance or proficiency for several reasons. First, Gollan et al. (2012) found that self-report more accurately estimated second-language competency than an objective measure. Furthermore, there is very little evidence regarding what objective measures to adopt when predicting non-native listeners' performance on an English auditory task. Finally, in past studies, we have observed that listener groups defined according to LEAP-Q measures showed reliable differences in performance tasks (e.g., Shi 2011, 2012; Shi & Morozova 2012; Shi & Zaki 2014). Such results provide evidence for the validity of LEAP-Q self-report measures for representing relevant differences in language background among listeners. See also Marian et al. (2007) for further evidence of the validity of the LEAP-Q.

The profile obtained via the LEAP-Q indicated that non-native listeners dominant in English were quite different from those dominant in other languages. A set of two-sample,

TABLE 1. Language profiles of the RED, RRD, and SSD listener groups

	RED	RRD	SSD
Age of testing (years)	27.07 ± 4.70	29.31 ± 4.13	30.00 ± 4.92
Age of acquisition (years)			
Listening/speaking	9.92 ± 4.70	12.85 ± 4.20	15.92 ± 6.87
Reading	10.54 ± 4.18	13.76 ± 4.42	16.77 ± 7.61
Length of immersion (years)			
Country	16.84 ± 3.19	10.78 ± 4.84	11.76 ± 3.97
Family	10.13 ± 8.59	3.05 ± 4.49	6.94 ± 8.12
School/work	14.35 ± 4.54	8.01 ± 4.40	11.67 ± 4.30
Self-rated proficiency in English (0–10)			
Listening	9.15 ± 0.90	8.08 ± 1.44	8.08 ± 1.19
Speaking	9.00 ± 0.82	6.62 ± 1.45	7.46 ± 1.45
Reading	9.15 ± 1.07	8.00 ± 1.68	8.00 ± 1.68

Data are expressed in the form of mean ± 1 standard deviations. Native listeners' age of testing was 27.31 ± 5.68 years. All variables refer to English. Proficiency ratings use a scale 0–10 with 0 being the least and 10 being the most. Specific descriptors are provided in the LEAP-Q (Marian et al. 2007).

RED, English-dominant Russian-English bilingual; RRD, Russian-dominant Russian-English bilingual; SSD, Spanish-dominant Spanish-English bilingual.

two-tailed Student's *t* tests was conducted to evaluate group differences in language experience profiles, controlling for type I error using a Bonferroni adjustment ($\alpha = 0.05/24 = 0.002$, as there were three pairwise comparisons for each of the eight language background variables). Compared with their RRD and SSD counterparts, RED listeners had significantly longer immersion in an English-speaking country (RED–RRD: $p = 0.001$; RED–SSD: $p = 0.001$), significantly more learning and use of English at school or work (RED–RRD: $p = 0.001$), and significantly higher English speaking proficiency (RED–RRD: $p < 0.001$; RED–SSD: $p = 0.002$). No significant differences were found for any variable between RRD and SSD groups.

Stimuli

Boothroyd and Nittrouer's (1988) sentences of HP, LP, and ZP were employed as the stimuli. There were 80 sentences belonging to three categories, depending on the amount of context: 20 HP sentences (e.g., "Cooks make hot food"), 20 LP sentences (e.g., "Feet catch slow thieves"), and 40 ZP sentences (e.g., "Tin bread blue more"). All sentences were composed of four meaningful monosyllabic content words. The same words were used in each sentence type; however, words were organized such that the HP sentences were semantically and syntactically intact, the LP sentences retained syntactic but not semantic cues, and the ZP sentences did not have either type of cue. Following Boothroyd and Nittrouer, these HP, LP, and ZP repertoires were divided into 4, 4, and 8 five-sentence lists, respectively.

Boothroyd and Nittrouer (1988) selected words according to the Thorndike and Lorge (1944) count of word frequency. Eighty percent of these words had a frequency above 10 per million and are therefore common words. To make sure that non-native listeners' performance was not due to their limited vocabulary, we asked low performers to identify words that appeared unfamiliar to them after the experiment. Low performers were defined as those who obtained a raw score <75% (i.e., >20 errors) when they were scored based on words in the HP conditions (i.e., with both semantic and syntactic cues present, a listener should be expected to recognize all four words if she/he knew these words). It turned out that only one RRD and three SSD listeners were low performers and the worst raw score was 58.75%. These listeners were presented a list of the words included in the experiment on paper and asked to circle words that they did not know. No words were singled out by any individual. All listeners claimed that they knew all the words presented to them.

All stimuli were rerecorded by a native speaker of American English who was a professional speaker with extensive expertise in American dialectology. The speaker read the materials, using no overt regional accent, in a sound-treated room at Haskins Laboratories using a head-mounted microphone (Audio-Technica ATM75, Tokyo, Japan), recorded to.wav using a Roland Edirol R09 24-bit recorder (Hamamatsu, Japan) with a 44.1 kHz sampling rate. Utterances were read from a printed list with the HP, LP, and ZP sentences randomized within block. Each of the sentence lists was produced twice. The second author and a third person, both native speakers of American English, listened to the two copies of each stimulus and selected the better copy (clear articulation, no distortion, etc.). Sentences were individually normalized in their root mean square amplitude to a 1000-Hz tone and edited to remove plosive burst distortions. All the above processing was performed using Pro Tools v7.3.1 (Avid Technology, Tewksbury, MA). In the end, eight unique randomizations of the sentence stimuli were made. The order of HP, LP, and ZP sentence lists was randomized across these eight versions of randomization, as done in Boothroyd and Nittrouer (1988).

Procedure

Following Boothroyd and Nittrouer (1988) and other previous studies, sentences were presented in a speech-spectrum steady-state noise, generated by a GSI-61 audiometer (Grason-Stadler, Madison, WI). Listeners were randomly presented with one randomization version of the stimuli binaurally at 65 dB SPL through a pair of supra-aural headphones (Telephonics, Farmingdale, NY). In the pilot phase of the study involving 21 listeners (11 EMN, 5 RED, and 5 RRD listeners), sentences were presented at 0 dB SNR. This SNR resulted in floor-level performance for ZP sentences in many non-native listeners. At the floor level, $p_i = 0$, $1 - p_i = 1$, and $\log_{10}1 = 0$, rendering the denominator meaningless in Equation (4). To address this problem, the SNR was subsequently changed to +3 dB SNR, which effectively avoided the floor-level performance and made it possible to calculate *k*.

All 21 listeners were invited to come back for an additional test of sentences after a minimum of 6 months. One EMN, 3 RED, and 4 RRD participants returned and were tested for sentences at +3 dB SNR. These listeners' performance at 0 dB SNR was discarded. Listeners recruited after the procedural change were only tested at +3 dB SNR. In summary, we only report findings from 4 groups of 13 listeners for +3 dB SNR.

Listeners were alerted at the beginning of each list to expect meaningful or nonsense sentences (Boothroyd & Nittrouer 1988; Benki 2003). They were instructed to verbally repeat every word in the sentence they heard as accurately as they could. Listeners were told that scoring was based on parts of each sentence, so they were encouraged to guess if not sure. For non-native listeners whose accent might pose a challenge for online scoring, verbal responses were recorded using a head-mounted condenser microphone (AKG C420) powered by a Crown Ph-1A phantom power supply, and digitized at 44.1 kHz into Audacity v1.2.6 (SoundForge.net) on a Compaq Presario laptop. These listeners were scored online and scoring was subsequently checked offline based on the recorded responses. Scoring of words was based on broad phonemic transcription with allowances made for frequent second-language patterns. For example, a tap or trill for English /r/ was not treated as a production error for Spanish listeners.

RESULTS

j Factor

Six sets of scores were obtained from each listener in the sentence recognition tests, depending on whether responses were scored for individual words or as a whole for HP, LP, or ZP sentences (Fig. 1). The overall pattern revealed that performance was better when sentences had cues (HP > LP > ZP) and performance was better with EMN than non-native listeners. EMN listeners' performance approached ceiling for the HP sentences, especially when scored by words. Recall that both semantic and syntactic cues were present in HP sentences.

Three sets of *j* factors (j_{ZP}, j_{LP}, and j_{HP}) were derived per listener group. Note that, while all EMN listeners had raw performance scores no higher than 95% correct based on the whole sentence, 8 of these listeners yielded scores >95% based on words in the sentence. As pointed out by Boothroyd and Nittrouer (1988), *j*- and *k*-factor estimates are sensitive to outliers; their solution was to exclude data <5% and >95%. Here, the sheer portion of these data (62%) led us to seek another method of controlling the influence of outliers while preserving the bulk of the data. Excluding so much data would have reduced statistical power and yielded an inaccurate characterization of differences between native and non-native listeners, underestimating group differences that can be relevant in clinical assessment of bilingual clients. Our solution was to exclude data from participants whose values were two standard deviations from the group average in computing *j* (and *k* as well). This procedure excluded three EMN listeners for HP, one listener for LP, and one listener for ZP sentences. Similarly, even though no non-native listeners performed >95% or <5% for any of the sentences, one RED listener's HP performance, one SSD listener's HP and LP performance, and one SSD listener's ZP performance were beyond two standard deviations from the group average and were consequently excluded from the analyses. The "Discussion" section revisits this methodological difference between the present study and that of Boothroyd and Nittrouer. Nevertheless, exclusion of these individuals did not affect the statistical power ($\beta > 0.8$).

Group-average *j*, after the removal of outliers, is shown in Figure 2 for each type of sentence. An average *j* close to 4.0 can be seen across groups for ZP sentences, where there was virtually no redundancy across the four words in the sentence. When syntax was made available (ZP compared with LP sentences), *j* decreased, and likewise when semantic cues were added (LP compared with HP sentences). A two-way mixed design ANOVA was conducted with listener group (EMN, RED, RRD, and SSD) as the between-subjects factor and sentence type (HP, LP, and ZP) as the within-subjects factor. Results revealed no significant main effect of listener group or interaction of listener group × sentence type (Table 2). However, the main effect of sentence type was significant. Post hoc pairwise comparisons for sentence type indicated that *j* was significantly lower with HP than LP sentences ($p = 0.008$), indicating use of semantic cues; with LP than ZP sentences ($p = 0.006$), indicating use of syntactic cues; and HP than ZP sentences ($p < 0.001$).

Taken together, the results indicated that syntactic and semantic cues were helpful to reduce the independence of words in the sentence. Each of these two contextual cues significantly

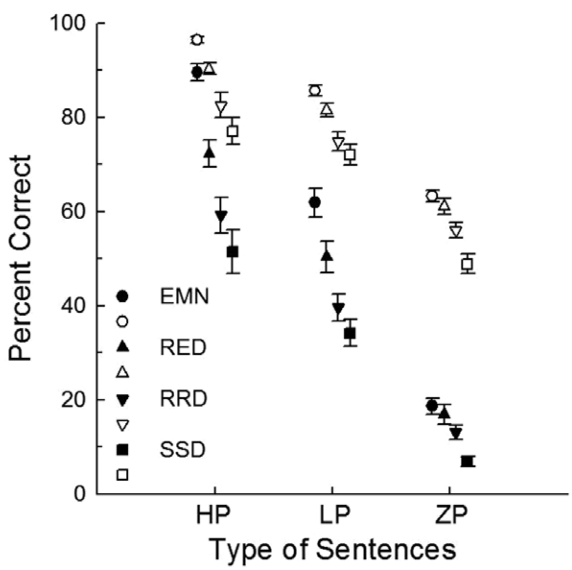

Fig. 1. Raw performance for HP, LP, and ZP sentences for EMN, RED, RRD, and SSD listener groups. *Unfilled and filled symbols* represent correct word and sentence scores, respectively. Error bars represent 1 standard error of the mean. EMN indicates English monolingual native; HP, high-predictability; LP, low-predictability; RED, English-dominant Russian-English bilingual; RRD, Russian-dominant Russian-English bilingual; SSD, Spanish-dominant Spanish-English bilingual; ZP, zero-predictability.

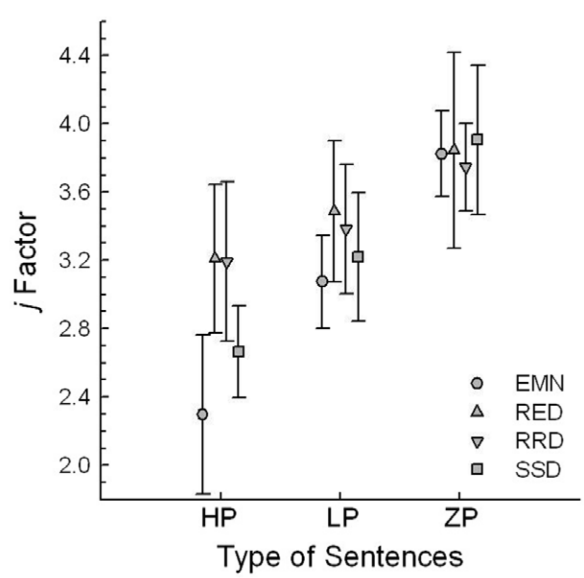

Fig. 2. The *j* factor for HP, LP, and ZP sentences for EMN, RED, RRD, and SSD listener groups. Error bars represent 1 standard error of the mean. EMN indicates English monolingual native; HP, high-predictability; LP, low-predictability; RED, English-dominant Russian-English bilingual; RRD, Russian-dominant Russian-English bilingual; SSD, Spanish-dominant Spanish-English bilingual; ZP, zero-predictability.

TABLE 2. Results of analysis of variance for *j* and *k* for words and sentences

Factors	df	F	p	η_p^2
j				
Sentence type	2, 78	19.611	**<0.001**	0.335
Listener group	3, 39	2.155	0.109	0.142
Sentence type × listener group	6, 78	1.286	0.273	0.090
k				
Context type	1.609, 67.584	188.096	**<0.001**	0.817
Listener group	3, 42	10.526	**<0.001**	0.429
Context type × listener group	4.827, 67.584	7.695	**<0.001**	0.355

Degrees of freedom reflect the exclusion of outliers. Statistics for the context type and context type × listener group for the k factor also reflects Greenhouse-Geisser adjustment. Shown in bold are statistically significant terms.

helped improve recognition of the sentence as a whole for all listeners, native and non-native. The magnitudes of these effects were comparable across non-native listeners, regardless of their language dominance and first language.

k Factor

To assess the relative weight of syntax and semantics in sentence recognition, three *k* factors were derived based on the raw scores in the sentence recognition tests (Fig. 3). The context effect due to syntax is denoted as k_s, obtained by comparing recognition scores of the LP versus ZP sentences. The context effect due to semantics is denoted as k_m, obtained by comparing recognition scores of the HP versus LP sentences. The combined effect of syntax and semantics is denoted as k_{s+m}, obtained by comparing recognition scores of the HP versus ZP sentences. As with *j*, outlier data were removed for the computation of *k*.

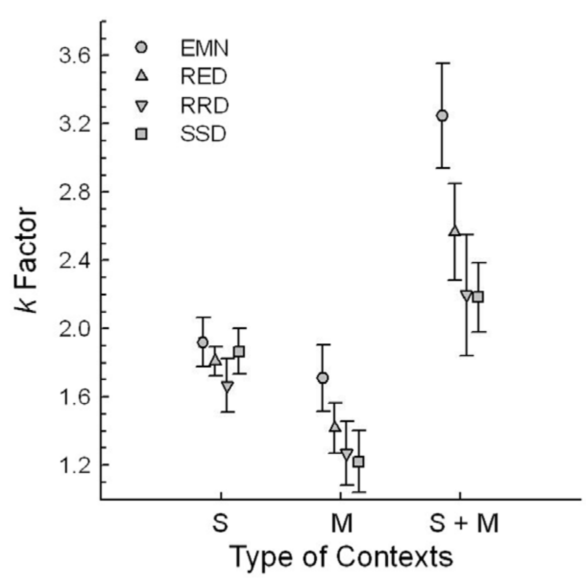

Fig. 3. The *k* factor for S, M, and S + M for EMN, RED, RRD, and SSD listener groups. Error bars represent 1 standard error of the mean. EMN indicates English monolingual native; M, semantics; RED, English-dominant Russian-English bilingual; RRD, Russian-dominant Russian-English bilingual; S, syntax; S + M, syntax + semantics; SSD, Spanish-dominant Spanish-English bilingual.

This procedure excluded one RRD listener for k_m, one EMN, one RRD, and one SSD listener for k_s, and one RRD and one SSD listener for k_{s+m}.

A two-way mixed-design ANOVA was conducted with listener group (EMN, RED, RRD, and SSD) as the between-subjects factor and context type (syntax, semantics, and syntax + semantics) as the within-subjects factor. Because the data failed Mauchly's test of sphericity ($W_2 = 0.757$, $p = 0.003$), degrees of freedom were adjusted with the Greenhouse-Geisser approach. Results revealed significant main effects of listener group and context type as well as a significant interaction of listener group × context type (Table 2), suggesting different cue weightings across listener groups. Table 3 illustrates post hoc pairwise comparisons of contextual cues by listener group. Across groups, k_{s+m} was significantly higher than k_m. For EMN and RED listeners, k_{s+m} was significantly higher than k_s, whereas this comparison was not significant for RRD and SSD listeners. For EMN listeners, k_s and k_m were similar, whereas k_s was significantly higher than k_m for RED, RRD, and SSD listeners. Taken in whole, syntax seemed to play a larger role than semantics for non-native listeners and semantics did not make significant contributions to sentence recognition for the two non-English-dominant groups.

Table 4 illustrates post hoc pairwise comparisons by type of context for the four listener groups. For syntax, no post hoc group-wise comparisons were statistically significant after Bonferroni correction ($\alpha = 0.05/18 = 0.003$). The only significant differences for semantics were between EMN and the two non-English-dominant groups, RRD and SSD. For syntax + semantics, EMN had a significantly higher k_{s+m} than all three non-native groups. No significant differences existed among the non-native groups.

Finally, k_m was compared against $k = 1$ using the one-sample Student's *t* test to see whether the "low" use of semantics by non-native listeners was different from "no" use at all (Shi 2014). The average of k_m was significantly higher than 1 for EMN ($t_{12} = 7.313$, $p < 0.001$) and RED ($t_{12} = 5.746$, $p < 0.001$) listeners, but not for the two non-English-dominant groups with Bonferroni's adjustment, $\alpha = 0.05/4 = 0.013$ (RRD:

TABLE 3. Post hoc pairwise comparisons among types of context per listener group for *k*

	k_m	k_s	k_{s+m}	Pairwise Comparisons
EMN	1.71	1.92	3.25	k_m vs. k_s: $p = 0.245$
				k_m vs. k_{s+m}: $p < 0.001$
				k_s vs. k_{s+m}: $p < 0.001$
RED	1.41	1.81	2.57	**k_m vs. k_s: $p = 0.001$**
				k_m vs. k_{s+m}: $p < 0.001$
				k_s vs. k_{s+m}: $p < 0.001$
RRD	1.27	1.67	2.20	**k_m vs. k_s: $p = 0.002$**
				k_m vs. k_{s+m}: $p < 0.001$
				k_s vs. k_{s+m}: $p = 0.019$
SSD	1.22	1.87	2.18	**k_m vs. k_s: $p = 0.004$**
				k_m vs. k_{s+m}: $p < 0.001$
				k_s vs. k_{s+m}: $p = 0.007$

Shown in bold are statistically significant differences after Bonferroni correction ($\alpha = 0.05/12 = 0.004$).
EMN, English monolingual native listeners; RED, English-dominant Russian-English bilingual listeners; RRD, Russian-dominant Russian-English bilingual listeners; SSD, Spanish-dominant Spanish-English bilingual listeners.

TABLE 4. Post hoc pairwise comparisons among listener groups per type of context for *k*

	EMN	RED	RRD	SSD
Semantics				
EMN		p = 0.025	**p = 0.003**	**p = 0.001**
RED			p = 0.218	p = 0.101
RRD				p = 0.706
SSD				
Syntax				
EMN		p = 0.200	p = 0.027	p = 0.597
RED			p = 0.117	p = 0.478
RRD				p = 0.065
SSD				
Semantics + syntax				
EMN		**p = 0.003**	**p < 0.001**	**p < 0.001**
RED			p = 0.117	p = 0.042
RRD				p = 0.948
SSD				

Shown in bold are statistically significant differences after Bonferroni correction (α = 0.05/18 = 0.003). EMN: English monolingual native listeners; RED: English-dominant Russian-English bilingual listeners; RRD: Russian-dominant Russian-English bilingual listeners; SSD: Spanish-dominant Spanish-English bilingual listeners.

EMN indicates English monolingual native; RED, English-dominant Russian-English bilingual; RRD, Russian-dominant Russian-English bilingual; SSD, Spanish-dominant Spanish-English bilingual.

t_{11} = 2.8978, p = 0.015; SSD: t_{12} = 2.439, p = 0.031). This analysis suggests that semantic skills were underdeveloped in non-native listeners for whom the test language was not their dominant language. All listeners, however, benefited from syntax in sentence recognition because k_s was significantly above 1 for all groups (EMN: t_{11} = 14.844, p < 0.001; RED: t_{12} = 18.604, p < 0.001; RRD: t_{11} = 8.477, p < 0.001; SSD: t_{11} = 13.005, p < 0.001). Because the data revealed that k_{s+m} was always greater than k_s for every listener, a t test was not needed to confirm that k_{s+m} was significantly above 1 as well.

Taken together, the results indicated that use of context was mediated by one's language background. Of the two types of cues, only native listeners weighted them equally, whereas syntax had a greater weight than semantics for all non-native listeners, showing a clear native-non-native difference in the weighting of contextual cues. Although English-dominant non-native listeners did not differ from native listeners in using either semantics or syntax to recognize sentences, they were significantly less effective than native listeners in jointly using the two contextual cues in the task.

DISCUSSION

This study assessed the effectiveness of using context to recognize English sentences in native EMN listeners as well as three groups of non-native listeners who differed in their native language and language dominance. Comparisons between EMN and the three bilingual groups quantify a native advantage in context use. Comparisons between RED and the two non-English-dominant groups evaluate differences due to language dominance. Comparisons between the RRD and SSD listeners show differences due to first language.

Before discussing our results and comparing them to previous studies, we would like to point out a key difference in analysis method between our study and Boothroyd and Nittrouer (1988). Boothroyd and Nittrouer's original study discarded extreme values using <5% and >95% as their criteria. This strategy is based on the consideration that extreme percent-correct scores have the potential to mathematically skew the log value. Due to the steep psychometric function of the test sentences, we were unable to avoid extreme values in EMN listeners, whose raw performance for the HP sentences was >95% in some cases, while keeping non-native listeners' performance above the floor level for the ZP sentences. As a result, we sought an alternative yet statistically based approach and excluded data whose j or k values were significantly different from the group average. This methodological decision allowed for greater preservation of data while minimizing the influence of extreme values. It also provided for a valid comparison of differences between native and non-native listeners' performance under the same SNR conditions, important for conducting assessment of this type in a clinical context, where performance is usually obtained across clients at same amplitude. Nevertheless, readers should take this methodological difference into account when comparing our j and k with Boothroyd and Nittrouer's values.

To reduce a ceiling effect, we could also test native listeners at a more stringent SNR than what was used for non-native listeners. This method, however, relies on the assumption that EMN listeners' use of context is constant regardless of SNR. This assumption has been challenged by a few recent studies (e.g., Bradlow & Alexander 2007; Caldwell & Nittrouer 2013; Shi 2014). For example, when Bradlow and Alexander compared the benefit due to clear speech between native and non-native listeners, they found that non-native listeners received more benefit than native listeners, a result quite unique in the context of abundant evidence that non-native listeners have difficulty making full use of the benefits of clear speech. This interesting result was, however, obtained by testing native listeners at a lower SNR than non-native listeners. The authors thus indicated that manipulation of test SNR could lead to very different conclusions. To provide a clear and valid estimate of differences among language groups in the same listening conditions, it was imperative that our study maintain the same test SNR across listener groups. It would be beneficial for future studies to investigate the SNR effect of context use in speech processing in non-native listeners. Another strategy could be to develop a new set of materials with a more gradual psychometric function than those developed by Boothroyd and Nittrouer (1988), thus avoiding extreme values while allowing for direct cross-group comparisons. This strategy, however, could complicate cross-study comparisons.

j Factor Across Listener Groups

Our first prediction was that j would show a pattern of HP < LP < ZP for all listeners, and that j values would be greater in non-native than native listeners. The results confirmed that j decreased in value as contextual cues became more abundant in the sentence (i.e., $j_{HP} < j_{LP} < j_{ZP}$). However, no significant intergroup differences were seen for j_{HP}, j_{LP}, or j_{ZP}. The current EMN group yielded j values within the 95% confidence intervals reported by Boothroyd and Nittrouer (1988) for all three types of sentences (cf. also Nittrouer & Boothroyd 1990, whose results were based on the same young adult participants as Boothroyd & Nittrouer 1988).

Figure 2 shows that j_{ZP} was statistically equivalent across listener groups, suggesting that, when no context cues were available, all listeners recognized parts of sentences in a comparable way. Because j_{LP} was also similar across listener groups and studies (Boothroyd & Nittrouer 1988; Nittrouer & Boothroyd 1990), the presence of syntax reduced recognition errors to a similar extent, regardless of language background. A caveat to this finding is that the four-word sentences used in these studies all had a fairly straightforward structure (see next section).

In contrast, while j_{HP} was largely comparable across groups (cf. the nonsignificant listener group effect in Table 2), Figure 2 does suggest higher j_{HP} for RED and RRD than for EMN listeners. Two-sample t tests comparing EMN with RED and RRD yield p values of 0.010 and 0.015, respectively, giving some support to the hypothesis that Russian native speakers, even those who consider themselves dominant in English, may make less effective use of context than native English listeners when sentences contain both semantic and syntactic cues. These differences, although small, invite future investigation.

k Factor Across Listener Groups and Cue Types

Our second prediction was that combined cues (k_{s+m}) would yield higher scores than individual cues (k_s and k_m), and that non-native listeners would show lower values for k. The results showed that the combination of syntax and semantics generally provided more information than each cue alone (i.e., $k_{s+m} > k_s$ and $k_{s+m} > k_m$), although differences between k_{s+m} and k_s were not statistically significant for the non-English-dominant groups (cf. Table 3). Whereas k_s was comparable across all four listener groups, k_{s+m} was lower in non-native than native listeners and k_m was lower in non-English-dominant non-native listeners (RRD and SSD) than EMN and RED listeners (cf. Table 4).

Two other studies obtained k values to assess the use of semantics and syntax alone and in combination (Boothroyd & Nittrouer 1988; Nittrouer & Boothroyd 1990). Both reported $k_{s+m} > k_m > k_s$, suggesting that the combination of syntax and semantics was most effective in enhancing sentence recognition. For example, Boothroyd and Nittrouer found that semantics increased performance by close to 100%, compared with 38% for syntax. Boothroyd and Nittrouer also found that the two cues in combination increased the accessibility of information by approximately 170%, when compared with no context.

These patterns differed in various ways from the present study, where k yielded the most complex effects due to the significant interaction between listener group and type of context. The EMN group in this work yielded a significantly higher k_{s+m} than Boothroyd and Nittrouer (1988), according to the group average and 95% confidence interval data provided in that study. In the current EMN group, the amount of accessible information increased performance by 225% (i.e., $k_{s+m} = 3.25$). It is unlikely that the inclusion of "extreme" data (HP scores >95% based on word components but not qualifying as outliers according to our ±2 SD criterion) could have artificially elevated the magnitude of k_{s+m}. Because HP scores based on words were involved in calculating both k_{s+m} and k_m, it would be odd that extreme scores should computationally have affected k_{s+m} but not k_m; in fact, k_m values were lower here than in Boothroyd and Nittrouer.

One potential cause of the cross-study difference other than the different methods of excluding extreme data could lie in the fact that a new recording of the stimulus materials was made for the present study. Studies of intelligibility have demonstrated extensive cross-speaker differences (Bradlow et al. 1996). Moreover, prosodic variables may have influenced listener performance. Although the speaker attempted to record all sentences in a similar intonation at the same pace (and our auditory impressions of the recordings suggested that he was successful in accomplishing this goal), some subtle differences in stress and phrasing could have existed among the three sentence types, perhaps increasing the intelligibility of the HP sentences more than expected from Boothroyd and Nittrouer's findings (1988). Possible differences in the nature of the recorded materials are factors that researchers should consider when making cross-study comparisons, but they do not interfere with main conclusions obtained here, which involve differences in cue weighting across listener groups. These group-wise differences were obtained using the same recording and are detailed in the following sections.

Relative Weighting of Semantics and Syntax in Non-Native Listeners

Past work made it difficult for specific predictions regarding the weighting of semantic and syntactic cues in non-native listeners (cf. prediction 3). The present results showed that non-native listeners made less effective use of semantics than syntax in recognizing sentences (i.e., $k_s > k_m$), a finding reminiscent of Mattys et al. (2010) and Shi (2014). In particular, the amount of information provided by semantics was not significant in non-English-dominant listeners (i.e., $k_{s+m} = k_s$), suggesting that these listeners had not developed skills to effectively use word meanings to predict upcoming words in the sentence or repair what had passed unrecognized.

Semantics • All three non-native groups under-used semantic context somewhat when compared with native listeners in Boothroyd and Nittrouer (1988), Olsen et al. (1997), and the present study. Semantics is a particularly difficult cue to acquire for non-native listeners who reported to be dominant in their first language (cf. Table 4). Kilborn (1992) found that German listeners who were highly proficient in English and had resided in the United States for many years still could not take advantage of semantic information to correctly recognize HP sentences in noise. These studies combined suggest that semantic use develops relatively slowly in non-native listeners.

Interestingly, RED listeners' k_m was significantly higher than 1, suggesting that, even though these non-native listeners might not reach native levels in semantic use, they did utilize the cue in the task. In contrast, for SSD and RRD listeners k_m was not significantly different from 1; also, as noted above, their k_s and k_{s+m} were comparable. These two findings both suggest that non-native listeners who relied more on their native language than on English were not using semantic cues effectively on this English sentence recognition test. This conclusion differs from that of Shi (2014), where non-native listeners as a group had k_m values significantly different from 1, implying that they could take advantage of semantics. One possible reason is that Shi included non-native listeners regardless of their language dominance. If all three non-native groups in the present study were combined, data from RED listeners could have brought up the average k_m for the entire group. A second reason could be that the sentences used in Shi contained richer semantic cues in the HP sentences (e.g., "The doctor prescribed the drug") and minimal semantics in the LP sentences (e.g., "She has known about the drug"). Such sentences may have provided greater variation in semantic information than the Boothroyd and Nittrouer (1988) stimuli used here.

Syntax • In contrast to semantics, the current non-native listeners' use of syntax was on a par with EMN listeners. EMN and RED listeners were comparable in their combined weighting of syntax and semantics ($k_{s+m} > k_m = k_s$), whereas for RRD and SSD listeners $k_{s+m} = k_s > k_m$. The pattern of the two non-native groups is consistent with the results of Oliver et al. (2012), who measured reaction time of German learners of Dutch to HP, LP, and ZP Dutch sentences presented in six-talker babble. Whereas German listeners improved in locating the target word in LP sentences over an intensive Dutch learning program, they could not utilize semantic cues to shorten the time to recognize target words in HP sentences. The investigators concluded that non-native listeners are better at exploiting syntactic than semantic cues.

The simple structure of the Boothroyd and Nittrouer (1988) sentences bears noting in the context of interpreting group-wise differences here. The Boothroyd and Nittrouer sentences each contained four words and usually consisted of a subject, verb, and object. The fourth word could be an adjective or adverb. As all sentences were declarative, listeners could use syntax as a cue to follow one auditory stream (Bregman 1990). A listener who missed one word could use syntactic knowledge to figure out, for example, that a verb was likely to follow a noun, even if the sentence did not make much sense. Such simple syntactic structures may place low demands on working memory and minimize group-wise differences. In a study of auditory stream segregation, Wu et al. (2011) proposed that spatial separation of speech and competing babble should permit stream segregation, relieving burdens on working memory. English and Mandarin listeners' recognition performance was obtained for simple syntactically correct but semantically nonsense sentences. The investigators found that spatial separation of sentences and masking babbles did not improve listeners' performance, suggesting that processing of simple sentences does not tax working memory to a significant extent.

Some authors have found that non-native listeners are less effective in syntactic processing than native listeners for complex sentences (Sanders et al. 2002; Sanders & Neville 2003). Nevertheless, comparable effectiveness in syntactic processing for native and non-native listeners has been reported in more complex tasks than employed in the present study. Rossi et al. (2006) reported that even non-native listeners with low proficiency in the second language processed syntax in a similar way as native listeners when word order was violated, although some group-wise differences were seen when listeners were presented sentences with violations in verb conjugation. Errors in word order versus verb conjugation could hence present different degrees of difficulty to non-native listeners. In the present study, syntax was reflected in violations in word order rather than in verb conjugation. Therefore, the finding of effective non-native use of syntax is in line with Rossi et al.'s conclusion.

Another point on the stimulus materials, acknowledged by Boothroyd and Nittrouer (1988), is that semantic cues were not completely removed from all LP sentences. A few LP sentences (e.g., "Cats get good ears.") are not necessarily improbable. Although such sentences were not numerous, their inclusion could have the effect of reducing the magnitude of k_m, downplaying the importance of semantics but increasing the magnitude of k_s and the evident importance of syntax. Finally, as indicated earlier, there may have been subtle prosodic differences among sentence types, for example, between LP and ZP sentences. The non-native listeners could have exploited sentential prosody when following the auditory steam, again contributing to increased magnitude of k_s. Conceivably, such effects could also be more extreme in non-native than native listeners.

Language Background

This study included non-native listeners of two first languages to ensure that findings were not limited to a single bilingual population. The present results showed no significant differences between Russian and Spanish native listeners for *j* or *k*, semantic or syntactic use, suggesting that first language was not a critical factor in cue weighting. In contrast, language dominance was a significant determinant of listeners' performance. Previous studies have found this variable to be effective in capturing performance differences among non-native listeners of various levels. Shi and Morozova (2012), for instance, showed that, although RED bilinguals (similar in profile to the RED listeners in the present study) misrecognized significantly more English words in quiet than EMNs, their phonemic errors followed the monolingual pattern. On the other hand, Russian-dominant listeners (similar to the RRD listeners here) not only made significantly more mistakes on the task but also did so in a way that could be explained by their first-language phonology. In combination, the results of these two studies imply that language dominance is an effective and comprehensive variable, helpful to researchers and clinicians who wish to predict a non-native listener's recognition of English speech.

While language dominance offers a convenient description of non-native listeners, it does not directly explain processes in language acquisition that could have led to the intergroup differences. To understand the possible causes for the differences, one may need to look at other language background variables. The differences in English learning, use, and proficiency between the RED and RRD or SSD groups were clear. RED listeners learned English earlier, had been exposed to English for a longer period of time, used English to a higher degree at home, and achieved greater English proficiency than the other two non-native groups. In the future, assessment of individual variables such as age of acquisition, length and intensity of exposure, etc., could provide further insight into differences in cue weighting and inform bilingual clinical work. A longitudinal strategy may prove especially fruitful in understanding whether or when semantic uses catch up with syntactic uses.

CONCLUSIONS

Regardless of their language dominance or first language, the listeners in this study used syntactic cues to a comparable extent as their native peers. However, a language-dominance effect was seen for semantic use: While English-dominant non-native listeners used semantics comparably to native listeners, non-native listeners who self-reported to be dominant in their first language demonstrated lower use of semantics than native listeners. In fact, semantics provided little additional recognition benefit to non-English-dominant non-native listeners. In addition, while native listeners applied comparable weight to syntax and semantics when recognizing sentences, non-native listeners relied on syntax more heavily than semantics. Finally, a native language advantage was found in the combined use of semantics and syntax, where native listeners outperformed all non-native listeners including those dominant in English. In combination, the data indicate that cue type and listener background must both be considered in understanding non-native English listeners' performance on English language tasks in the presence of noise.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the time and effort of all the volunteers who participated in this study, the assistance of Doug Honorof and Yvonne Law in stimulus preparation, and the helpful comments of Doug Whalen and Anders Lövquist on an earlier version of the manuscript.

Portions of this study were presented at the 160th Acoustical Society of America in Cancun, Mexico, in 2010.

The authors declare no conflicts of interest.

Address for correspondence: Lu-Feng Shi, Department of Communication Sciences and Disorders, Long Island University–Brooklyn Campus, Brooklyn, NY 11201, USA. E-mail: lu.shi@liu.edu

Received February 12, 2015; accepted December 4, 2015.

REFERENCES

Akker, E., & Cutler, A. (2003). Prosodic cues to semantic structure in native and nonnative listening. *Bilingualism: Language and Cognition*, *6*, 81–96.

American National Standards Institute. (2010). ANSI-S3.6-2010, Specifications for Audiometers. New York: American National Standards Institute.

Aydelott, J., & Bates, E. (2004). Effects of acoustic distortion and semantic context on lexical access. *Language and Cognitive Processes*, *19*, 29–56.

Bell, T. S., & Wilson, R. H. (2001). Sentence recognition materials based on frequency of word use and lexical confusability. *J Am Acad Audiol*, *12*, 514–522.

Benki, J. R. (2003). Quantitative evaluation of lexical status, word frequency, and neighborhood density as context effects in spoken word recognition. *J Acoust Soc Am*, *113*, 1689–1705.

Benoit, C. (1990). An intelligibility test using semantically unpredictable sentences: Towards the quantification of linguistic complexity. *Speech Communication*, *9*, 293–304.

Boothroyd, A., & Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *J Acoust Soc Am*, *84*, 101–114.

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *J Acoust Soc Am*, *121*, 2339–2349.

Bradlow, A. R., Torretta, G. M., Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Commun*, *20*, 255–272.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: Bradford.

Bronkhorst, A. W., Bosman, A. J., Smoorenburg, G. F. (1994). A model for context effects in speech recognition. *J Acoust Soc Am*, *93*, 499–509.

Caldwell, A., & Nittrouer, S. (2013). Speech perception in noise by children with cochlear implants. *J Speech Lang Hear Res*, *56*, 13–30.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*, 3–42.

Cutler, A., Mehler, J., Norris, D., et al. (1986). The syllable's differing role in the segmentation of French and English. *J Memory Lang*, *25*, 385–400.

Eisenberg, L. S., Shannon, R. V., Martinez, A. S., et al. (2000). Speech recognition with reduced spectral cues as a function of age. *J Acoust Soc Am*, *107*(5 Pt 1), 2704–2710.

Golestani, N., Rosen, S., Scott, S. K. (2009). Native-language benefit for understanding speech-in-noise: The contribution of semantics. *Biling (Camb Engl)*, *12*, 385–392.

Gollan, T. H., Weissberger, G. H., Runnqvist, E., et al. (2012). Self-ratings of spoken language dominance: A multi-lingual naming test (MINT) and preliminary norms for young and aging Spanish-English Bilinguals. *Biling (Camb Engl)*, *15*, 594–615.

Kalikow, D. N., Stevens, K. N., Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *J Acoust Soc Am*, *61*, 1337–1351.

Kilborn, K. (1992). On-line integration of grammatical information in a second language. In R. Harris (Ed.), *Cognitive Processing in Bilinguals* (pp. 337–350). Amsterdam, The Netherlands: Elsevier Science Publishers.

Marian, V., Blumenfeld, H. K., Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *J Speech Lang Hear Res*, *50*, 940–967.

Mattys, S. L., White, L., Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *J Exp Psychol Gen*, *134*, 477–500.

Mattys, S. L., Carroll, L. M., Li, C. K. W., et al. (2010). Effects of energetic and informational masking on speech segmentation by native and non-native speakers. *Speech Commun*, *52*, 887–899.

Mayo, L. H., Florentine, M., Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *J Speech Lang Hear Res*, *40*, 686–693.

McClelland, J. L., Mirman, D., Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends Cogn Sci*, *10*, 363–369.

Nakatani, L. H., & Dukes, K. D. (1973). A sensitive test of speech communication quality. *J Acoust Soc Am*, *53*, 1083–1092.

Nittrouer, S., & Boothroyd, A. (1990). Context effects on phoneme and word recognition by young children and older adults. *J Acoust Soc Am*, *84*, 101–114.

Nittrouer, S., & Lowenstein, J. H. (2010). Learning to perceptually organize speech signals in native fashion. *J Acoust Soc Am*, *127*, 1624–1635.

Oliver, G., Gullberg, M., Hellwig, F., et al. (2012). Acquiring L2 sentence comprehension: A longitudinal study of word monitoring in noise. *Bilingualism: Lang Cogn*, *15*, 841–857.

Olsen, W. O., Van Tasell, D. J., Speaks, C. E. (1997). The Carhart memorial lecture, American auditory society, Salt Lake City, Utah, 1996. Phoneme and word recognition for words in isolation and in sentences. *Ear Hear*, *18*, 175–188.

Rossi, S., Gugler, M. F., Friederici, A. D., et al. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *J Cogn Neurosci*, *18*, 2030–2048.

Sanders, L. D., & Neville, H. J. (2003). An ERP study of continuous speech processing. II. Segmentation, semantics, and syntax in non-native speakers. *Brain Res Cogn Brain Res*, *15*, 214–227.

Sanders, L. D., Neville, H. J., Woldorff, M. G. (2002). Speech segmentation by native and non-native speakers: The use of lexical, syntactic, and stress-pattern cues. *J Speech Lang Hear Res*, *45*, 519–530.

Shi, L. F. (2010). Perception of acoustically degraded sentences in bilingual listeners who differ in age of english acquisition. *J Speech Lang Hear Res*, *53*, 821–835.

Shi, L. F. (2011). How "proficient" is proficient? Subjective proficiency as a predictor of bilingual listeners' recognition of English words. *Am J Audiol*, *20*, 19–32.

Shi, L. F. (2012). Contribution of linguistic variables to bilingual listeners' perception of degraded English sentences. *J Speech Lang Hear Res*, *55*, 219–234.

Shi, L. F. (2014). Measuring effectiveness of semantic cues in degraded English sentences in non-native listeners. *Int J Audiol*, *53*, 30–39.

Shi, L. F., & Morozova, N. (2012). Understanding native Russian listeners' errors on an English word recognition test: Model-based analysis of phoneme confusion. *Int J Audiol*, *51*, 597–605.

Shi, L. F., & Zaki, N. A. (2014). Psychometric function for NU-6 word recognition in noise: Effects of first language and dominant language. *Ear Hear*, *35*, 236–245.

Speaks, C., & Jerger, J. (1965). Method for measurement of speech identification. *J Speech Hear Res*, *8*, 185–194.

Thorndike, E., & Lorge, I. (1944). *The Teacher's Word Book of 30000 Words*. New York, NY: Columbia University Press.

van Wijngaarden, S. J., Bronkhorst, A. W., Houtgast, T., et al. (2004). Using the speech transmission index for predicting non-native speech intelligibility. *J Acoust Soc Am*, *115*, 1281–1291.

Versfeld, N. J., Daalder, L., Festen, J. M., et al. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *J Acoust Soc Am*, *107*, 1671–1684.

Wu, X., Yang, Z., Huang, Y., et al. (2011). Cross-language differences in informational masking of speech by speech: English versus Mandarin Chinese. *J Speech Lang Hear Res*, *54*, 1506–1524.